



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

# THE STATA JOURNAL

## **Editor**

H. Joseph Newton  
Department of Statistics  
Texas A&M University  
College Station, Texas 77843  
979-845-8817; fax 979-845-6077  
jnewton@stata-journal.com

## **Editor**

Nicholas J. Cox  
Department of Geography  
Durham University  
South Road  
Durham DH1 3LE UK  
n.j.cox@stata-journal.com

## **Associate Editors**

Christopher F. Baum  
Boston College

Nathaniel Beck  
New York University

Rino Bellocco  
Karolinska Institutet, Sweden, and  
University of Milano-Bicocca, Italy

Maarten L. Buis  
Tübingen University, Germany

A. Colin Cameron  
University of California–Davis

Mario A. Cleves  
Univ. of Arkansas for Medical Sciences

William D. Dupont  
Vanderbilt University

David Epstein  
Columbia University

Allan Gregory  
Queen's University

James Hardin  
University of South Carolina

Ben Jann  
University of Bern, Switzerland

Stephen Jenkins  
London School of Economics and  
Political Science

Ulrich Kohler  
WZB, Berlin

Frauke Kreuter  
University of Maryland–College Park

Peter A. Lachenbruch  
Oregon State University

Jens Lauritsen  
Odense University Hospital

Stanley Lemeshow  
Ohio State University

J. Scott Long  
Indiana University

Roger Newson  
Imperial College, London

Austin Nichols  
Urban Institute, Washington DC

Marcello Pagano  
Harvard School of Public Health

Sophia Rabe-Hesketh  
University of California–Berkeley

J. Patrick Royston  
MRC Clinical Trials Unit, London

Philip Ryan  
University of Adelaide

Mark E. Schaffer  
Heriot-Watt University, Edinburgh

Jeroen Weesie  
Utrecht University

Nicholas J. G. Winter  
University of Virginia

Jeffrey Wooldridge  
Michigan State University

**Stata Press Editorial Manager**  
**Stata Press Copy Editors**

Lisa Gilmore  
Fred Iacoletti and Deirdre Skaggs

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- Current Contents/Social and Behavioral Sciences®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)
- Scopus™
- Social Sciences Citation Index®

**Copyright Statement:** The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata, Mata, NetCourse, and Stata Press are registered trademarks of StataCorp LP.

# Importing financial data

Mehmet F. Dicle  
Loyola University New Orleans  
New Orleans, LA  
mfdicle@loyno.edu

John Levendis  
Loyola University New Orleans  
New Orleans, LA  
jlevendi@loyno.edu

**Abstract.** In this article, we describe two commands—`fetchyahooquotes` and `fetchyahookeystats`—that import historical financial data and key current financial statistics from Yahoo! Finance.

**Keywords:** dm0061, `fetchyahooquotes`, `fetchyahookeystats`, finance, financial data, stocks, exchange-traded funds, historical data, Yahoo! Finance

## 1 Introduction

Yahoo! Finance offers the public large amounts of financial and economic data. Often, small amounts of these data can be imported into Stata without much difficulty. However, to use these data in econometric analysis, an automated download procedure may prove useful. We will describe two commands—`fetchyahooquotes` and `fetchyahookeystats`—that automate the process of importing financial data from Yahoo! Finance.

Many universities do not have subscriptions to costly finance databases. Though Yahoo! Finance is not a substitute for these databases, it is a useful alternative. `fetchyahooquotes` and `fetchyahookeystats` make access to financial data fast and easy. These two commands are useful for instructors who work with financial data during their lectures and need quick access to current data. They are especially useful for finance project assignments because students do not have to spend long hours downloading data into Excel spreadsheets and merging them into portfolios. These commands are also important for researchers and investors.

The `fetchyahooquotes` command is used to download time series of the more common financial statistics of multiple financial instruments. For example, one could download the daily opening and closing prices over the past 10 years of IBM and Microsoft stocks. `fetchyahookeystats` is similar to `fetchyahooquotes` except that it downloads only the current day's key financial statistics for multiple financial instruments (for example, today's opening and closing prices for IBM and Microsoft).

## 2 The fetchyahooquotes and fetchyahookeystats commands

### 2.1 Syntax

```
fetchyahooquotes namelist, freq(d|w|m|v) [field(string) chg(string)  
      save(filename) start(date) end(date) ff3]
```

```
fetchyahookeystats namelist, field(string) [save(filename)]
```

*namelist* is a list of ticker symbols to be downloaded from Yahoo! Finance's application programming interface. Ticker symbols are separated by spaces. The ticker symbol will become part of the variable name. Regardless of whether the symbol includes special characters, the ticker is typed (including the special character) exactly as it is used by Yahoo! Finance's application programming interface. Any special characters in the ticker symbol will be replaced with underscores (`_`) in the variable name because symbols are not allowed in Stata variable names.

### 2.2 Options for fetchyahooquotes

freq(*d|w|m|v*) specifies the frequency of the historic price: daily (*d*), weekly (*w*), monthly (*m*), or dividends only (*v*). If the frequency is either daily, weekly, or monthly, then the variable with the symbol name is the adjusted closing price. If the frequency is to include only the dividends, then the symbol name is the dividend payment. freq() is required.

field(*string*) specifies variables to download along with the adjusted close and the date. These variables can be the following: open (*o*), high (*h*), low (*l*), close (*c*), and volume (*v*).

chg(*string*) is the periodic return. Three different periodic changes can be calculated for the adjusted closing price: natural log difference (*ln*), percentage change (*per*), and symmetrical percentage change (*sper*). The change is based on the continuous trading assumption. Thus, although the *tsset* is the date (which may contain gaps), returns are calculated assuming there are no gaps in the data.

save(*string*) is the output filename. A Stata data file is created in the current working directory.

start(*date*) is the starting date for the prices in day-month-year format (for example, 1mar2010).

end(*date*) is the ending date for the prices in day-month-year format (for example, 23feb2011).

`ff3` specifies that Fama/French daily factors are downloaded from “Kenneth R. French—Data Library”.<sup>1</sup>

## 2.3 Options for `fetchyahookeystats`

`field(string)` is Yahoo! Finance’s specific field code corresponding to a key statistic. `field()` is required. These field codes include (but are not limited to) the following:

a	Ask	l1	Last trade price
b	Bid	m3	50-day moving average
b4	Book value	m4	200-day moving average
c	Percent change	n	Name
c1	Change	o	Open
d	Dividend per share	p	Previous close
d1	Last trade date	p5	Price/sales
e	Earning per share	p6	Price/book
f6	Float shares	q	Ex-dividend date
g	Day’s low	r	Price/earnings
h	Day’s high	s	Symbol
j	52-week low	s7	Short ratio
j1	Market capitalization	v	Volume
j4	Earnings before interest, taxes, depreciation, and amortization	x	Exchange
k	52-week high	y	Dividend yield

`save(filename)` is the output filename. A Stata data file is created in the current working directory.

## 3 Using `fetchyahooquotes` to import historical prices

### ▷ Example

*Single company historic share prices.* In this example, we use `fetchyahooquotes` to import the adjusted daily closing price of Microsoft shares from the beginning of 2010 to the end of 2010. Fama/French daily factors and the Microsoft share prices are downloaded. The command also calculates the log difference change of the daily closing share prices.

```
. fetchyahooquotes MSFT, freq(d) chg(ln) start(01jan2010) end(31dec2010) ff3
Fama/French daily factors are downloaded from 'Kenneth R. French - Data
> Library'.
MSFT is downloaded.
    time variable:  date, 04jan2010 to 31dec2010, but with gaps
                delta:  1 day
```

---

1. Available at [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
date	252	18446.69	104.8328	18266	18627
adjclose_M-T	252	26.26262	2.06914	22.33	30.32
ln_MSFT	251	-.0003306	.013845	-.0419884	.0515297
ff3_Mkt_RF	252	.0007242	.0117564	-.0408	.0453
ff3_SMB	252	.000479	.00558	-.0202	.0165
ff3_HML	252	-.0000512	.0051936	-.0162	.0154
ff3_RF	252	7.94e-07	2.71e-06	0	.00001

By specifying the `chg(ln)` option, we can calculate the log difference changes in these prices. With the `ff3` option, Fama/French factors are also downloaded. Fama/French original variable names are used with a prefix of “`ff3_`”.<sup>2</sup>

### ► Example

*Multiple companies’ historic share prices.* In this example, we download monthly historical data about the closing share prices of Microsoft and IBM, and we compute the log difference changes. The monthly data are actually the data for the first day of trading in the month.

```
. fetchyahooquotes MSFT IBM XYZ, freq(m) chg(ln)
MSFT is downloaded.
IBM is downloaded.
XYZ does not have sufficient number of observations.
      time variable:  date, 02jan1962 to 03oct2011, but with gaps
              delta:  1 day

. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
date	599	9815.922	5254.355	732	18903
adjclose_M-T	308	14.50955	11.7808	.08	45.87
ln_MSFT	307	.0189928	.1042121	-.4207915	.430783
adjclose_IBM	598	34.29679	40.42496	1.53	185.88
ln_IBM	597	.0072513	.0705309	-.3042277	.3029297

```
. list in 1/3
```

	date	adjclo~T	ln_MSFT	adjclo~M	ln_IBM
1.	02jan1962	.	.	2.45	.
2.	01feb1962	.	.	2.43	-.0081968
3.	01mar1962	.	.	2.41	-.0082645

Missing values are reported because Microsoft did not exist in 1962. No data could be downloaded for XYZ because that symbol does not exist.

2. Explanations of these variables are available at  
[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data\\_Library/f-f\\_factors.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html).

## ▷ Example

Multiple companies' indices and a foreign symbol's historic values. We requested to download the daily high, low, opening, and adjusted closing prices of the following: IBM, Google, the S&P 500 (denoted as ^GSPC), and BMW (noted as BMW.DE, BMW shares that are listed in the XETRA in Germany).

```
. fetchyahooquotes IBM GOOG ^GSPC BMW.DE, freq(d) field(h l o)
IBM is downloaded.
GOOG is downloaded.
^GSPC is downloaded.
BMW.DE is downloaded.
    time variable:  date, 03jan1950 to 27oct2011, but with gaps
                  delta:  1 day

. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
date	15628	7696.559	6532.211	-3650	18927
open_IBM	12544	191.2321	138.0621	41	649
high_IBM	12544	192.861	138.8806	41.75	649.88
low_IBM	12544	189.7026	137.3624	40.63	645.5
adjclose_IBM	12544	34.20004	40.19985	1.38	190.53
open_GOOG	1813	443.0304	132.7804	99.19	741.13
high_GOOG	1813	448.0714	133.3038	101.74	747.24
low_GOOG	1813	437.4407	131.7214	95.96	725
adjclose_G~G	1813	442.7003	132.486	100.01	741.79
open__GSPC	15557	399.931	461.7358	16.66	1564.98
high__GSPC	15557	402.633	464.93	16.66	1576.09
low__GSPC	15557	397.0871	458.2907	16.66	1555.46
adjclose__~C	15557	400.0182	461.7963	16.66	1565.15
open_BMW_DE	2281	38.26399	9.994679	17.28	73.77
high_BMW_DE	2281	38.76526	10.09087	17.82	73.85
low_BMW_DE	2281	37.73375	9.891829	16	71.57
adjclose_B~E	2281	36.32488	10.28628	16.36	73.52

Because variable names in Stata cannot have special characters, such as “.”, and because our convention is to include the ticker name in the variable names, this presents a problem. We would like to report the daily open price of BMW.DE as a variable named `open_BMW.DE`; however, Stata does not allow periods in variable names. Thus `fetchyahooquotes` replaces special characters with an underscore: `open_BMW.DE` becomes `open_BMW_DE`. Likewise, `open_^GSPC` becomes `open__GSPC`.



### ▷ Example

*Multiple companies' historic dividends.* In this example, we download the dividend payments of IBM, BMW, and Ford.

```
. fetchyahooquotes IBM BMW.DE F, freq(v)
IBM is downloaded.
BMW.DE does not have sufficient number of observations.
F is downloaded.
      time variable:  date, 06feb1962 to 08aug2011, but with gaps
              delta:  1 day

. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
date	312	10583.16	4671.158	767	18847
dividends_~M	196	.1837389	.1535859	.001	.75
dividends_F	116	.2428889	1.113203	.01907	12.06693

The dividend payments of IBM and Ford are downloaded, but there are not enough observations in the dataset to download dividend payments of BMW.

## 4 Using fetchyahookeystats to import historical prices

### ▷ Example

*Using fetchyahookeystats to import key statistics for multiple companies.* This example uses `fetchyahookeystats` to download today's (28 October 2011) key statistics for IBM, Google, BMW, and the S&P 500.

```
. fetchyahookeystats IBM GOOG ^GSPC BMW.DE,
> field(n s l1 a b d1 g h k e n s k e L) save(my_portfolio)
(17 vars, 4 obs)
```

Under the `field()` option, we specified that we wanted the full name of the item (**n**), the ticker symbol (**s**), the last traded price (**l1**), the ask price (**a**), the bid (**b**), the last trade date (**d1**), the day's low (**g**), the day's high (**h**), the 52-week high (**k**), and the earnings per share (**e**). Note that, **n**, **s**, **k**, and **e** are included twice in the `field()` option. The `field()` option also includes **L**, which is not listed as a known Yahoo! Finance key statistics field.

```
. describe
Contains data from my_portfolio.dta
  obs:      4
  vars:      11                      28 Oct 2011 09:00
  size:      308
```

variable name	storage type	display format	value label	variable label
Symbol	str6	%9s		
Name	str17	%17s		
Last_Trade_Pr-e	float	%9.0g		
Ask	double	%10.0g		
Bid	double	%10.0g		
Days_Low	float	%9.0g		
Days_High	float	%9.0g		
_52_Week_High	float	%9.0g		
Earnings_per_-e	double	%10.0g		
Unknown_Field_1	str6	%9s		
Last_Trade_Date	double	%td		

Sorted by:

The `field()` qualifiers included twice (`n`, `s`, `k`, and `e`) are ignored for their second occurrence. The statistic associated with the `L` qualifier is downloaded and named `Unknown_Field_1`.

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Symbol	0				
Name	0				
Last_Trade	4	531.678	549.1867	61.65	1280.48
Ask	2	330.79	380.5932	61.67	599.91
Bid	2	330.54	380.2962	61.63	599.45
Days_Low	4	529.35	548.0494	60.86	1277.01
Days_High	4	534.0375	550.9103	62.47	1284.39
_52_Week_High	4	568.255	588.8249	73.85	1370.58
Earnings_per	3	16.59033	11.31233	7.745	29.337
Unknown_Field_1	0				
Last_Trade	4	18928	0	18928	18928

### About the authors

Mehmet F. Dicle is an assistant professor of finance at Loyola University New Orleans.

John Levendis is an assistant professor of economics at Loyola University New Orleans.